

What is claimed is:

1 1. A method of dynamically deploying services in a computing network, comprising steps of:
2 receiving client requests for a selected service;
3 serving the received requests from a first server when the selected service has not yet been
4 dynamically deployed;
5 effecting a dynamic deployment by programmatically moving the selected service from the
6 first server to one or more other servers when the dynamic deployment is triggered; and
7 serving the received requests from the one or more other servers after the effecting step
8 causes the selected service to be dynamically deployed.

1 2. The method according to Claim 1, further comprising the steps of:
2 monitoring a number of the received client requests for the selected service; and
3 triggering the dynamic deployment when the monitored number exceeds a predetermined
4 threshold.

1 3. The method according to Claim 2, wherein the predetermined threshold applies to a
2 plurality of dynamically deployable services.

1 4. The method according to Claim 2, wherein the predetermined threshold applies to the
2 selected service.

1 5. The method according to Claim 2, wherein a value of the predetermined threshold applies

to all of the one or more other servers.

6. The method according to Claim 2, wherein values of the predetermined threshold apply to individual ones of the one or more other servers.

7. The method according to Claim 2, wherein a value of the predetermined threshold is specified by a systems administrator.

8. The method according to Claim 2, wherein a value of the predetermined threshold is specified as a default value.

9. The method according to Claim 2, wherein a value of the predetermined threshold is specified programmatically.

10. The method according to Claim 2, wherein the monitoring step counts the received client requests at individual ones of the one or more other servers.

11. The method according to Claim 2, wherein the monitoring step counts the received client requests at a plurality of the one or more other servers.

12. The method according to Claim 1, further comprising the steps of:
monitoring a load on the computing network; and

3 triggering the dynamic deployment when the monitored load exceeds a predetermined
4 threshold.

1 13. The method according to Claim 1, wherein the programmatically moving step further
2 comprises the step of issuing a deployment request for the selected service.

1 14. The method according to Claim 13, wherein the deployment request comprises a service
2 description of the selected service encoded in a standardized service description notation.

1 15. The method according to Claim 14, wherein the service description comprises an interface
2 definition of a dynamic deployment service and an implementation definition of the dynamic
3 deployment service.

1 16. The method according to Claim 15, wherein the dynamic deployment service resides on
2 the first server.

1 17. The method according to Claim 13, further comprising steps of:
2 receiving the deployment request at a particular one of the one or more other servers, the
3 particular one being that server to which the selected service is being dynamically deployed;
4 issuing a subsequent deployment request, responsive to the receiving step, from the
5 particular one to the first server; and
6 receiving, responsive to the subsequent deployment request, a deployment response from

the first server.

18. The method according to Claim 17, wherein the subsequent deployment request comprises a SOAP ("Simple Object Access Protocol") request.

19. The method according to Claim 17, wherein the subsequent deployment request comprises an XML ("Extensible Markup Language") Protocol request.

20. The method according to Claim 17, wherein the subsequent deployment request identifies the selected service.

21. The method according to Claim 17, wherein the subsequent deployment request provides information about run-time conditions on the particular one.

22. The method according to Claim 17, wherein the deployment response comprises executable code to be deployed on the particular one.

23. The method according to Claim 21, wherein the deployment response comprises executable code which is adapted to the run-time conditions on the particular one.

24. The method according to Claim 22, further comprising the step of deploying the executable code on the particular one.

1 25. The method according to Claim 1, further comprising the steps of:
2 transparently routing the received client requests using a repository which tracks whether
3 the selected service is deployed on the one or more other servers; and
4 wherein the serving steps serve the received requests from the first server or from the one
5 or more other servers, depending on the transparently routing step.

1 26. A system for dynamically deploying services in a computing network, comprising:
2 means for receiving client requests for a selected service;
3 means for serving the received requests from a first server when the selected service has
4 not yet been dynamically deployed;
5 means for effecting a dynamic deployment by programmatically moving the selected
6 service from the first server to one or more other servers when the dynamic deployment is
7 triggered; and
8 means for serving the received requests from the one or more other servers after the
9 effecting step causes the selected service to be dynamically deployed.

1 27. A computer program product for dynamically deploying services in a computing network,
2 the computer program product embodied on one or more computer-readable media and
3 comprising
4 computer-readable program code means for receiving client requests for a selected
5 service;

computer-readable program code means for serving the received requests from a first server when the selected service has not yet been dynamically deployed;

computer-readable program code means for effecting a dynamic deployment by programmatically moving the selected service from the first server to one or more other servers when the dynamic deployment is triggered; and

computer-readable program code means for serving the received requests from the one or more other servers after the effecting step causes the selected service to be dynamically deployed.